Computers & Geosciences 54 (2013) 142–147

Contents lists available at SciVerse ScienceDirect

Computers & Geosciences

journal homepage: www.elsevier.com/locate/cageo

# Improving bathymetric images exploration: A data mining approach

Luis Fernando Planella Gonzalez, Maria Alejandra Gomez Pivel, Duncan Dubugras Alcoba Ruiz *

*Business Intelligence Research Group, Computer Science Graduate Program, Faculty of Informatics, Pontifical Catholic University of Rio Grande do Sul - PUCRS, Av. Ipiranga, 6681 Predio 32, sala 628, 90619-900, Porto Alegre, RS, Brazil*

ARTICLE INFO

*Article history:*
Received 8 November 2012
Received in revised form
28 December 2012
Accepted 29 December 2012
Available online 16 January 2013

*Keywords:*
Data mining
Bathymetry
Image processing
Cold-water corals

ABSTRACT

Bathymetry is the science of measuring and charting the depths to determine the topography of the seafloor and other bodies of water. It has several important practical and academic applications. For this reason, having computational tools capable of analyzing bathymetric charts would be useful for domain experts studying the various problems related to water depth. Data mining is a well known technique for extracting information from large datasets, but cannot be directly applied to images. The contribution of this work is an approach for using data mining in bathymetry images. We propose a method for processing input images, in order to extract records and their features, which can be processed by classic data mining algorithms. Additionally, we also propose techniques to visualize both data mining results and map characteristics. For evaluation purposes, the proposed approach was applied to a cold-water corals dataset, in order to predict where corals are likely to be found, under a domain expert supervision.



## 1. Introduction

Bathymetric maps have several important applications, like exploration and exploitation of mineral resources, navigation planning and study of deep water circulation and sediment transport (Brown et al., 2007). Bathymetry maps can cover the entire extent of the Earth globe, and they can show subtle variations between nearby areas, as well as similar areas which are distant from each other. Considering those points, it would be interesting to have available computational tools which could aid domain experts on the analysis of problems on various fields related to bathymetry.

The expected operation for such a tool is that a domain expert can manually label map areas, or alternatively, import labeled areas from a known database. The available labels (classes) are arbitrary, according to the specific problem being addressed. Then, assuming that the knowledge is limited to certain specific areas, the computational tool should, from the map characteristics on labeled areas, autonomously label other areas, where the class label was previously unknown. It might be the case that only a few areas have known data, and the computational tool should still infer the other area's labels with an acceptable performance.

### 1.1. Data mining

We propose the usage of data mining as the computational tool previously referred. Data mining is a process for automatic information discovery in large databases (Tan et al., 2005), and is capable of performing descriptive and predictive tasks (Han et al., 2005). Descriptive tasks characterize the general properties of the data, while predictive tasks perform inference on the current data, in order to make predictions.

### 1.2. Classification

The data mining task which fulfills the requirements described above is classification, which is a predictive task that infers the value of a nominal class attribute, given the other attributes on the dataset. A learning algorithm is used to analyze the dataset (called training set), inducting a classification model. That model is then applied to a test dataset, inferring the value for the class attribute on each record. There are several classification models, each having multiple well known algorithms implementing them. This work employs decision trees, as they are widely used and well known, but most classification algorithms should be suitable as well.

### 1.3. Decision trees

Decision trees are tree data structures where each intermediary node contains nova tests over data set attributes, and leaf nodes

* Corresponding author. Tel.: +55 51 3320 3611; fax: +55 51 3320 3621.
*E-mail addresses:* luisfpg@gmail.com (L.F.P. Gonzalez), maria.pivel@pucrs.br (M.A.G. Pivel), duncan.ruiz@pucrs.br, duncanruiz@gmail.com (D.D.A. Ruiz).

0098-3004/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cageo.2012.12.009

contains the inferred class value. Fig. 1 shows an example of a decision tree. Some of the classical decision trees algorithms are *CART* (Breiman et al., 1984) and *C4.5* (Quinlan, 1993). There are also algorithms which are called ensemble methods, aggregating the predictions of multiple decision trees in order to increase the prediction performance. A notable example of an ensemble method is the *Random Forests* algorithm (Breiman, 2001).

### 1.4. Classification evaluation

Classification evaluation is normally done by splitting the training set into training and test sets (Tan et al., 2005). That way, the inferred model can be evaluated, as the class labels of the training set are previously known. The plain splitting of the training set is called holdout. When the holdout method is repeated multiple times, averaging the precisions of each iteration, that is called random subsampling. However, a more controlled method for evaluation is the $n$-fold cross validation, where $n$ is the number of subsets the original data set is splitted into. At each iteration, one of the subsets is used as test, and the others, as training. The process is repeated $n$ times, and the estimated precision is the average of each iteration. The evaluation results can be summarized as a precision, which is calculated by dividing the number of correctly classified records by the total number of records (and can be calculated for an individual class), or as the confusion matrix, which presents all class labels in both rows and columns. The cell $ij$ contains the number of records which where predicted to belong to class $i$, and were known to belong to class $j$. Hence, the main diagonal (where $i=j$) contains the correctly classified records, according to the dataset, while other cells contains classification errors.

### 1.5. Image data mining

Classical data mining algorithms work with tabular data-collections of records (rows) and their attributes (features/columns). Therefore, in order to apply those algorithms over other data types, as it is the case with bathymetric images, a preprocessing must be performed, transforming the original data into a tabular form. This is done by extracting both records and attributes from the original images.

The first aspect which should be decided is the granularity of a record. Generally, when the purpose of data mining is to label or cluster an image database, each image is mapped to a single record (Ding et al., 2009; Kitamoto, 2002). Otherwise, when multiple data is expected to be processed on each image, it is segmented into multiple records. A straightforward approach is to split images into a fixed-size grid (Gueguen and Datcu, 2007). Alternatively, computer vision techniques can be used to delimit salient objects, being each object a distinct record (Fan et al., 2008).

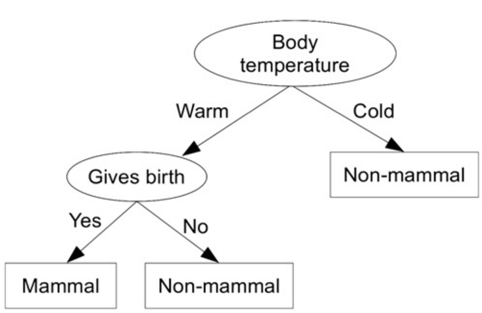


**Fig. 1.** Example of a decision tree with two classes.
*Source*: Tan et al. (2005).

For attribute extraction from image contents, image processing techniques are used. Some of the well-known image processing techniques are as follows:

- *Pixel value statistics*: Statistical variables, like mean, standard deviation, variance or mode, over pixel values. Each variable is used as a data mining attribute.
- *Color histogram* (Novak and Shafer, 1992): A histogram consisting of $n$ bins. Each bin corresponds to a discretization of the pixel values, and is used as an attribute. The discretization can be done in several ways. An example is to convert the image color model into a limited palette, being each distinct color a bin. Another example is to divide the pixel luminosity by the number of bins, resulting in the destination bin for that pixel.
- *Wavelet coefficients* (Mallat, 1989): Wavelet coefficients can be used to capture shapes, textures and locations. For images, the Discrete Wavelet Transform (DWT) may be used. It takes as input a square image having size of a power of 2, and generates four components, each with half of the original image size: approximation, horizontal details, vertical details and diagonal details. The approximation is a downsampled image, and the others contain coefficients which can be used to reconstruct the original image. This is the level 1. The DWT can then be applied again to the approximation component, generating a smaller approximation, with the respective coefficients, increasing the level. To generate data mining attributes, the DWT can be used up to a given number of levels, using all coefficients from all components as attributes.

Image processing tasks are usually highly dependent on the problem domain. So, not only the parameters and variations for those techniques can fit best distinct problem domains, but also other techniques, more specific to the target domain can be used.

## 2. Proposed method

The method proposed in this work for employing data mining to analyze bathymetry images consists in defining the expected format for input images and how to extract records and attributes from them. Additionally, we propose a method to visualize results and map characteristics.

### 2.1. Expected image format

It would be extremely complex to handle any possible bathymetry image, due to distinct cartographic projections, scales, angles, color models and any other element which can be present on such images. So, in order to effectively extract data mining records from images, a straightforward approach is to define an expected image format.

In the proposed method, the input image is expected to be on the geographic (or Platte Carrée, or equirectangular) projection, as it allows processing the entire coordinates space. Also, it considers meridians and parallels to be perpendicular in relation to each other, making it a natural choice for square sub-images, as described later in the text. One thing to note, however, is that the geographic projection distorts the map near the poles. The method does not take those distortions into account, but they can impact on the data mining results. A possible way to mitigate this problem is to discard records in extreme latitudes when performing data mining tasks.

The image should be in the gray scale color model, as a single data (the bathymetry at each pixel) needs to be read. In order to differentiate land from water, it is assumed that pixels which represent land are pure black (RGB values are all 0). Other values represent distinct depths, being lighter values shallow areas, and darker values, deep areas.

A bathymetric image with the aforementioned characteristics is available from NASA's Visible Earth project, at http://visibleearth.nasa.gov/view_rec.php?id=8392, as shown in Fig. 2.

### 2.2. Records extraction

We propose to split the image in squared sub-images of arbitrary size, each one dubbed *tile*. Each tile will be treated as a separate record for data mining.

As this work is focused on bathymetry, and purely black pixels represent dry land, tiles containing those should be ignored. This is important because pixels with value zero represent a major gap in relation to the shallower valid value, which is 255, hence, interfering on calculations.

### 2.3. Attribute extraction

For attribute extraction, all the well-known techniques described in Section 1.5 are used, and others are proposed.

The well-known techniques which were employed, as well as their semantics for the bathymetry domain are:

- *Pixel value statistics*: Two attributes are used: mean and standard deviation. Mean captures the average depth, while standard deviation describes how much the depth varies along the tile. For example, continental shelves tend to have high means and low standard deviations; deep plateaus, to have both mean and standard deviation with low values; while continental slopes will have high standard deviations.
- *Color histogram*: A $n$-bin histogram, where the target bin for each given pixel with value $v$ (as the expected image format is in gray scale, the only available measure is the value) is calculated by $round\_floor(256/n)$. Each bin value is used as a data mining attribute. The histogram allows capturing tiles where specific depth ranges are important for the classification.
- *Wavelet coefficients*: Two input variables are expected: a tile base size $b$ and the number of DWT levels $l$. The original tile size is arbitrary, but the DWT requires the image to be squared, having each dimension as a power of 2. To extract attributes, the tile is resized to $b \times b$, then the DWT is applied $l$ times in the approximation component. Each of the resulting coefficients is used as a data mining attribute. For example, for $b=128$ and $l=4$, there will be 256 attributes. Wavelet coefficients capture low-level image features, like textures, sizes and positions.

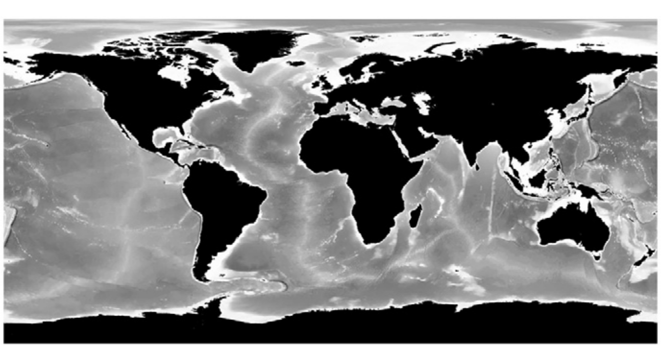

**Fig. 2.** Bathymetry map used in this work, available from NASA's Visible Earth project.

Besides the characteristics captured by those attributes, we consider the approximate tile shape to be important for the classification. Therefore, we propose two new attribute sets:

- *Regions*: Regions capture the approximate tile shape, disregarding the average depth. This technique consists in splitting the tile in a sub-areas grid, where each sub-area is named *region*. An input parameter $r$ is used to determine the region sizes, being the tile splitted in $r \times r$ regions. Each region consists in a delta between the pixel value means of both the region and the entire tile, and generates one attribute on the resulting data set. Fig. 3 depicts the values for each region in a tile. In the example, lighter (shallower) regions will have higher values, and the darkest (deepest) regions will have negative values.
  As similar shapes may appear in arbitrary rotations, and regions are square areas, only rotations multiple of 90° can be captured. So, as an heuristic, the highest valued region among the four corners is placed on the upper-left corner, either by rotating (90°, 180° or 270°) or mirroring (horizontally and/or vertically). However, intermediate angles might not be recognized by regions.
- *Vector*: This attribute group consists of a vector capturing the dominant tile direction and intensity. Vectors are calculated over regions, which are means, and not over individual pixels. This is done to prevent outliers from interfering too much on the vector. Being min and max the lowest and highest-valued regions, the vector has the following characteristics:

$$
\begin{aligned}
&x = x_{max} - x_{min}\\
&y = y_{max} - y_{min}\\
&z = value_{max} - value_{min}\\
&angle_{xy} = a\tan(\tfrac{y}{x})\\
&angle_{zy} = a\tan(\tfrac{y}{z})\\
&angle = \sqrt{x^2 + y^2 + z^2}
\end{aligned}
$$

For the purpose of capturing the dominant tile direction and intensity, the vector coordinates ($x$, $y$ and $z$) are not needed. Fig. 4 shows the vector on the coordinates system, and a vector overlaid on the regions that generated it.

For both regions and vector, a threshold may be defined, for tiles to be considered totally flat. If that threshold is not reached,

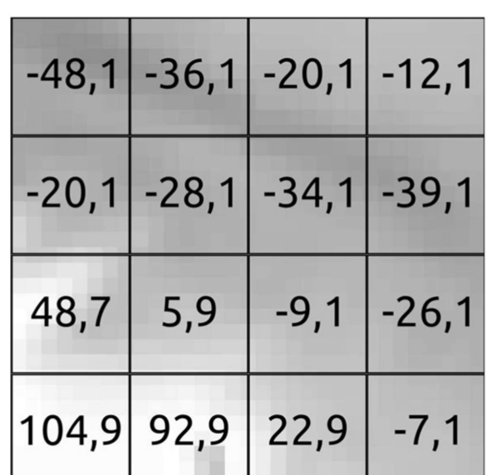


**Fig. 3.** Region attribute values overlaid in a tile.

the tile is considered to be totally flat, which means that all attributes in both regions and vector are passed to the data mining algorithm with value zero.

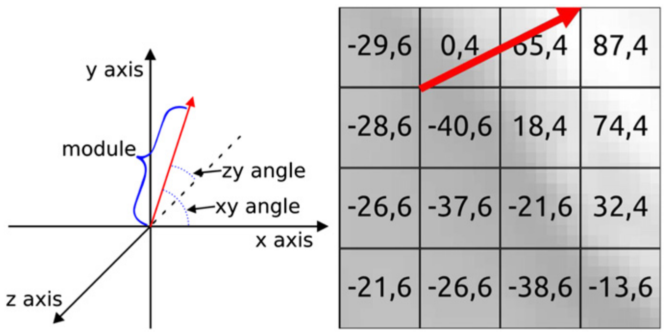


**Fig. 4.** Vector on the coordinates system and positioned over regions.

### 2.4. Results and map visualization

An important aspect to help the domain expert in terms of understanding the data mining results, as well as the map characteristics is visualization. A straightforward way to represent the distinct classes is by using distinct colors for them. It is important, however, to differentiate the hard truth data and the one which has been inferred by data mining. A possible solution for that is to vary the color opacity in both cases, which will lead to more saturated tiles when manually labeled, and lighter tiles when inferred. Fig. 5 shows examples of tile coloring for two classes, showing both initially labeled and inferred label tiles.

Not only the classification results can be visually represented on the map, but also the extracted attributes from each tile. From the attributes presented in Section 2.3, one of the most useful to help understanding map characteristics is the vector, as it captures the tile dominant direction in a succinct way. Vectors are also naturally represented as arrows, so, the corresponding vector can be rendered over each tile. Both modulus and *xy* angle are easily represented using two dimensions. The *zy* angle, however, is represented according to the discretization of its values to four different levels. Each level will vary both thickness and arrow head size. Fig. 6 shows several possible cases for vectors, including all four levels of *zy* angles, distinct rotations and modules, and also two highlighted areas where no vectors were drawn, because they were assumed to be flat (according to the flat tile threshold described in Section 2.3).

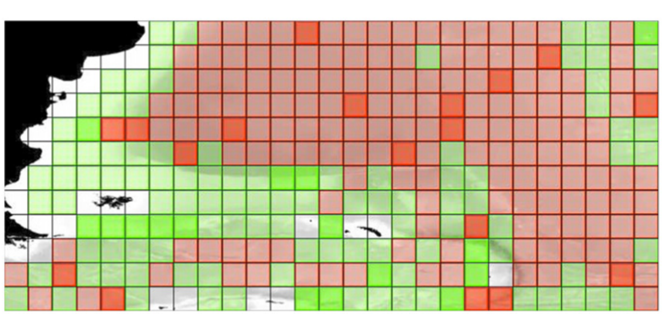

**Fig. 5.** Example of map coloring using two classes, showing tiles which were initially labeled (more saturated) and tiles with inferred labels (less saturated). (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

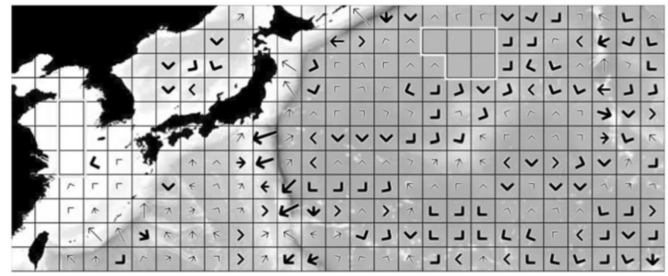

**Fig. 6.** Vectors overlaid on tiles. The four distinct levels of *zy* angles, distinct rotations on the *xy* angle and distinct modules are shown. Also highlighted are the two areas without vectors, as those tiles were considered to be totally flat.

## 3. Method evaluation

In order to evaluate the proposed approach of data mining in bathymetric images, we decided to infer the global distribution of deep-sea corals using a database of their well-known geographical distribution (*i.e.* based on sampling records). On one hand, deep-sea corals were chosen because their distribution is closely conditioned by geology, since it determines the type of substrate and relief, which in turn influence ocean currents. Furthermore, corals were chosen because they play an important role as habitat providers for fisheries. In fact, the knowledge of the distribution and structure of the deep-water coral related ecosystems has become relevant after the decline of coastal fisheries. Additionally, deep-sea corals constitute important high-resolution paleoceanographic records (Sherwood and Risk, 2007) because they incorporate into their skeletons different proportions of trace elements and their isotopes that vary along their growth in consonance with climatic and oceanographic changes.

Despite having been discovered in the XVIII century, only recently the deep-water corals received considerable interest and research upon them has significantly advanced. Given their importance, it is fundamental to locate and map habitat areas to complement existing information on their distribution, to understand patterns of occurrence around the world, and to provide location and extent information towards protection of reef areas from damaging activities (Freiwald et al., 2004). Since the actual survey of the whole sea floor is a virtually impossible task, data mining could prove itself as a useful tool to address their study.

The distribution database used was UNEPs World Conservation Monitoring Centre, Cold-Water Corals, Version 2.0 (Rogers and Hall-Spencer, 2005) downloaded from the Ocean Biogeographic Information System (http://www.iobis.org). The database comprises 6553 records obtained from 1869 up to 2005, containing the geographic coordinates and corals species sampled. The records are geographically scattered, reflecting both the patchy distribution of corals and the bias of data due to the non-uniform sampling effort (Fig. 7). That figure was generated on http://iobis.org/mapper/, by choosing the "Cold Water Corals" data set and selecting "Points" on the "Layer" menu.

The objective of data mining in this particular case is to find areas which potentially contain corals, among those which had no information on the original dataset. In this case, no information could mean that either no corals exist or that the area was not (correctly) sampled. Given that enclosed seas and lakes are significantly different from open seas, these areas were excluded from the analysis by painting them with absolute black color in the bathymetry map. As a result, areas such as the Caspian and Black seas were ignored in the image processing, just as if they were solid land.

The original database records could not be used directly on the proposed method, because classic classification algorithms handle a single class per record, and multiple coral species could be mapped to the same tile. To overcome this, we have defined two classes, "Yes" and "No", representing the presence and absence of corals in a given tile. Only tiles on the latitude range present on the database were considered, in order to avoid projection distortions on unsampled areas near the poles, which could interfere in results. Each record was mapped to a tile, according to its coordinates. Tiles which contain at least one coral sample were labeled as "Yes". We also need to label other tiles as "No", despite that the original data set has no information about the "No" class.

Besides labeling tiles as "No", the tile size must also be defined. In order to accomplish both, a random subset of yet unlabeled tiles was labeled "No". The number of tiles chosen was the same of "Yes" tiles, in order to achieve class balance on the training set. That training set was then used on a 10-fold cross validation, and the precision for class "Yes" was stored. This process was repeated 10 times for each of the following tile sizes: $4\times4$, $8\times8$, $16\times16$, $32\times32$, $48\times48$ and $64\times64$. Each subset was as disjunct as possible at each iteration (tiles were only repeated after all others had been used). Table 1 presents the average "Yes" precision for executions on each of the mentioned tile sizes, indicating that $4\times4$ is the tile size with greater precision. The Random Forests algorithm was chosen, as it generally provides better classification performance than other single tree algorithms.

The final step is to label tiles without any corresponding records on the original database. As it is composed solely of positive records, the result is also focused on the "Yes" class. At each data mining iteration, not only the 10-fold cross validation was executed, but also the inferred model was used to classify the unlabeled tiles. The set of tiles assigned to the "Yes" class was stored at each iteration. For the final result, a threshold was then used to consider as "Yes" tiles which were positively labeled by at least that given number of intermediate models. This way, the result will be robust in relation to noise originated from the random subsets of tiles labeled as "No". Table 2 shows the number of records labeled as "Yes" having the threshold varying from 1 to 10.

Finally, we considered the threshold of 10, as it still provides a considerable number of areas which potentially have corals, and provides the higher possible confidence. The result is depicted in Fig. 8.

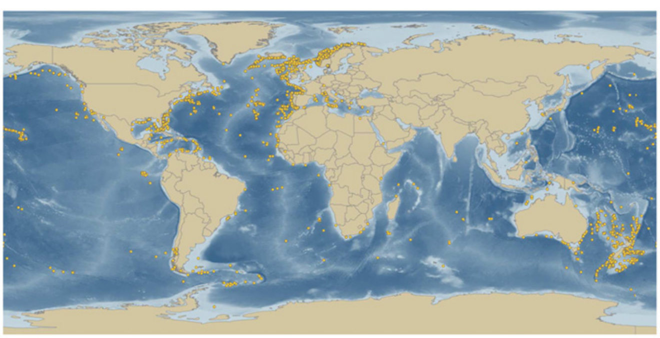

**Fig. 7.** Distribution of deep-water coral records. Each dot represents one record on the data set.

**Table 1**
Average precisions of 10 iterations consisting on 10-fold cross validation for each tile size.

| Tile size | "Yes" precision (%) |
|---|---|
| $4\times4$ | 90.5 |
| $8\times8$ | 88.1 |
| $16\times16$ | 84.3 |
| $32\times32$ | 79.5 |
| $48\times48$ | 74.8 |
| $64\times64$ | 72.1 |

**Table 2**
Number of tiles classified as "Yes" for each threshold.

| Threshold | "Yes" count |
|---|---|
| 1 | 144,939 |
| 2 | 119,581 |
| 3 | 105,107 |
| 4 | 95,573 |
| 5 | 88,142 |
| 6 | 81,958 |
| 7 | 75,754 |
| 8 | 68,840 |
| 9 | 58,265 |
| 10 | 44,231 |

## 4. Discussion

Deep-sea corals are more abundant in submarine mounds and along the slopes of continents and islands (Roberts et al., 2006). This pattern appears very well represented in the predicted distribution (Fig. 8).

Comparing the proven distribution (*i.e.* the distribution known from real sampling) and the expected distribution (*i.e.* based on data mining) we see that the latter is consistent both in terms of geographical location and relative to the seabed relief. However, compared to the proven distribution, the predicted distribution shows a more uniform pattern and suggests a more widespread occurrence of deep-sea corals. On the one hand, the homogeneity of the distribution of predicted occurrences (in comparison with documented occurrences) is easily understandable because of the heterogeneity of the oceanographic sampling effort. The North Atlantic is by far the most well studied and densely sampled basin in the world, and the concentration of the sampling effort in this basin certainly biases the census data. In fact, despite being the largest ecosystem of the world, covering around 60% of the solid surface of the Earth, the deep-sea is still poorly known compared to other marine ecosystems (Glover and Smith, 2003). Thus, distribution data are usually strongly biased and the real biogeography of deep-water species is still very limited.

At a first glimpse, the more widespread predicted occurrence of deep-sea corals shown in this study may seem to overestimate

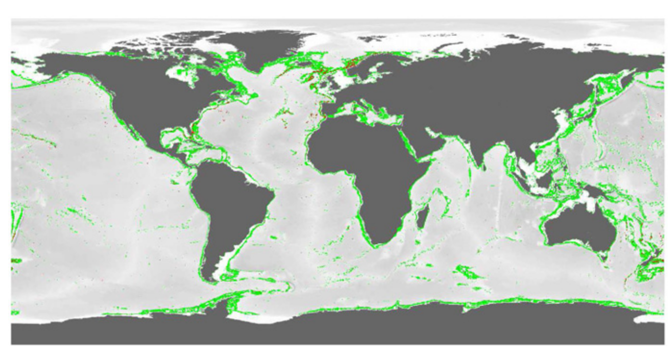

**Fig. 8.** Results from the experiment using a cell size of 4 pixels. Positive real values are in red, and predicted positive values are in light green. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

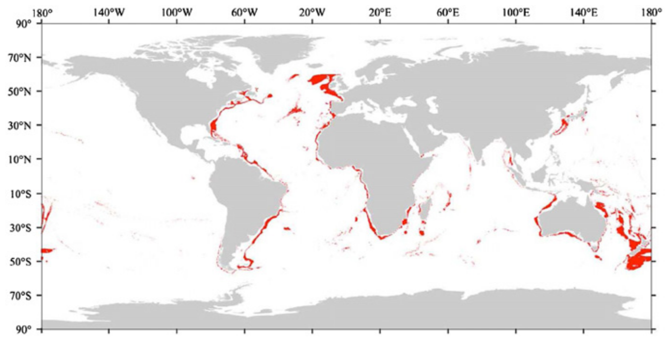


**Fig. 9.** Davies and Guinotte's predicted presence map for scleractinian framework-forming corals. White background indicates that these species are not likely to be found, red indicates probable presence. See original paper for details (Davies and Guinotte, 2011). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the areas suitable for deep-sea coral distribution. However, the fact of a grid cell being considered a likely candidate for the presence of corals does not mean that we assume that corals would be densely present on the seafloor in that grid cell. Even if the environmental conditions are suitable, the distribution of deep-sea corals is patchy.

Although geology and physiography may be considered the key parameters determining deep-sea coral distribution, other physical, chemical and biological oceanography parameters may also exert influence. Indeed, other studies (*e.g.* Davies et al., 2008; Tittensor et al., 2009; Davies and Guinotte, 2011) have estimated coral distribution using more comprehensive approaches including not only bathymetry but also different physicochemical and biological parameters such as primary productivity in surface waters, which represents a food source for the deep-sea. Nevertheless, the results obtained in this study based solely in bathymetry are comparable to those obtained in these more comprehensive studies. For example, Davies and Guinotte (2011) predicted the presence of the five main species of framework-forming scleractinian corals which together account for 50% of all the occurrences in UNEP's database (Rogers and Hall-Spencer, 2005). Since the geographical distribution of the other 50% of the occurrences practically covers the same areas, we can confidently compare our study to Davies and Guinotte's. Although their results are in higher spatial resolution, they show a similar pattern to the results obtained in our study (Fig. 9). The main difference between both studies is in high latitudes and at the northeastern Pacific, along the western margin of North America where Davies and Guinotte do not predict the occurrence of corals. However, at least in the case of the northeastern Pacific, the results obtained in our study appear to be consistent since the occurrence of deep-sea corals has already been documented for that region. For high latitudes, care must be taken since the results are very likely affected by projection distortions as mentioned in Section 2.1.

The similar results obtained between our study and Davies and Guinotte (2011) suggest that seabed relief is in fact the main driver of deep-sea coral distribution and that – at least for the study of deep-sea corals – a data mining approach using only this environmental parameter should provide reliable results. We believe that the method could be improved by combining other databases in order to provide answers to other geoscience related problems.

## 5. Conclusions

With the presented techniques, a data set compatible to data mining algorithms can be obtained from a bathymetry image with expected characteristics, by extracting records and attributes from it. With that data set, it is possible to apply several data mining tasks, like classification, regression, clustering and so on. For the scope of this work, classification was explored, as the motivation was that a domain expert could label specific areas on the map and have the computational tool to autonomously label the other areas.

The proposed method was tested with a cold-water corals data set, but any other problem related to bathymetry, which could be mapped to the expected format (one label per map tile) could be processed by data mining as well. The results which were obtained are consistent with the expected coral distribution (in terms of geology), and are comparable with another study (Davies and Guinotte, 2011), which used several other variables (than bathymetry) to predict its results.

Therefore, one of the biggest contributions of the proposed technique is to allow discovering information about partially known data having solely high resolution bathymetry imagery (freely available on the Internet) and the initially labeled data set. The results obtained could then receive a more thorough analysis using other, more specific approaches.

## References

Breiman, L., 2001. Random forests. Machine Learning 45 (1), 5–32.
Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., 1984. Classification and Regression Trees, 1st ed. Chapman and Hall/CRC, Belmont, CA, USA.
Brown, E., Coiling, A., Park, D., Phillips, J., Rothery, D., Wright, J., 2007. The Ocean Basins: Their Structure and Evolution, 2nd ed. Butterworth–Heinemann in Association with The Open University, Milton Keynes, England.
Davies, A.J., Guinotte, J.M., 2011. Global habitat suitability for framework-forming cold-water corals. PLoS ONE 6 (4), 5555.
Davies, A.J., Wisshak, M., Orr, J.C., Roberts, J.M., 2008. Predicting suitable habitat for the cold-water coral lophelia pertusa (scleractinia). Deep Sea Research Part I: Oceanographic Research Papers 55 (8), 1048–1062.
Ding, G., Wang, J., Xu, N., Zhang, L., 2009. Automatic image annotations by mining web image data. In: ICDM Workshops, pp. 152–157.
Fan, J., Gao, Y., Luo, H., Jain, R., 2008. Mining multilevel image semantics via hierarchical classification. IEEE Transactions on Multimedia 10 (2), 167–187.
Freiwald, A., Fossa, J., Grehan, A., Koslow, T., Roberts, J., 2004. Cold-Water Coral Reefs: Out of Sight – No Longer Out of Mind. UNEP-WCMC, Cambridge, England.
Glover, A.G., Smith, C.R., 2003. The deep-sea floor ecosystem: current status and prospects of anthropogenic change by the year 2025. Environmental Conservation 30 (03), 219–241.
Gueguen, L., Datcu, M., 2007. Image time-series data mining based on the information-bottleneck principle. IEEE Transactions on Geoscience and Remote Sensing 45 (4), 827–838.
Han, J., Kamber, M., Pei, J., 2005. Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers Inc, San Francisco, CA, USA.
Kitamoto, A., 2002. Spatio-temporal data mining for typhoon image collection. Journal of Intelligent Information Systems 19 (1), 25–41.
Mallat, S.G., 1989. A theory for multiresolution signal decomposition: the wavelet representation. IEEE Transactions on Pattern Analysis and Machine Intelligence 11 (7), 674–693.
Novak, C., Shafer, S., 1992. Anatomy of a color histogram. In: Proceedings CVPR '92, 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1992, pp. 599–605.
Quinlan, J.R., 1993. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers Inc, San Francisco, CA, USA.
Roberts, J.M., Wheeler, A.J., Freiwald, A., 2006. Reefs of the deep: the biology and geology of cold-water coral ecosystems. Science 312 (5773), 543–547.
Rogers, A., Hall-Spencer, J., 2005. Cold-Water Corals: Version 2.0. UNEP World Conservation Monitoring Centre (UNEP-WCMC).
Sherwood, O.A., Risk, M.J., 2007. Chapter 12: Deep-sea corals: new insights to paleoceanography. In: Marcel, C.H., Vernal, A.D. (Eds.), Proxies in Late Cenozoic Paleoceanography of Developments in Marine Geology, vol. 1. Elsevier, pp. 491–522.
Tan, P., Steinbach, M., Kumar, V., 2005. Introduction to Data Mining. Addison-Wesley Longman Publishing Co. Inc., Boston, MA, USA.
Tittensor, D.P., Baco, A.R., Brewin, P.E., Clark, M.R., Consalvey, M., Hall-Spencer, J., Rowden, A.A., Schlacher, T., Stocks, K.I., Rogers, A.D., 2009. Predicting global habitat suitability for stony corals on seamounts. Journal of Biogeography 36 (6), 1111–1128.